



Construct validity and assessment of the learning curve for the SIMENDO endoscopic simulator

E. G. G. Verdaasdonk,^{1,2} L. P. S. Stassen,^{1,2} M. P. Schijven,³ J. Dankelman¹

¹ Department of BioMechanical Engineering, Man Machine Systems Group, Delft University of Technology, Faculty of Mechanical, Maritime and Materials Engineering, Mekelweg 2, 2628 CD Delft, The Netherlands

² Department of Surgery, Reinier de Graaf Gasthuis, Delft, The Netherlands

³ Department of Surgery, University Medical Centre Utrecht, Utrecht, The Netherlands

Received: 21 September 2006/Accepted: 1 October 2006/Online publication: 16 February 2007

Abstract

Background: The SIMENDO is an affordable virtual reality simulator designed to train basic psychomotor skills for endoscopic surgery. This study aimed first to establish construct validity by determining which parameters can discriminate groups with different experience levels, and second to establish the extent to which training is useful by determining when inexperienced groups reach expert level.

Methods: The study participants were divided into four groups according to their experience with endoscopic procedures: experienced group (group A, > 50 procedures performed, $n = 15$), intermediate group (group B, 1–50 procedures performed, $n = 18$), endoscope navigation group (group C, endoscope navigation experience, $n = 14$), and novice group (group D, no endoscopic experience, $n = 14$). Each participant performed three repetitions of six consecutive exercises. The parameters studied were task time, path length of the instruments, and number of errors (collisions). Some participants continued training up to 10 repetitions to get insight in the learning curve.

Results: Group A (expert) outperformed all the other groups (B, C, and D) in terms of total median task time ($p < 0.05$), groups C and D in terms of path length, and group D in terms of collision frequency in the first two repetitions. Group B (intermediate) outperformed group D (novice) in total time and endoscope path length for all repetitions, and group C (camera navigation) outperformed group D (novice) in the first repetition. Less experienced groups D and C did not reach expert level for the task time within 10 repetitions, and group B reached it after the eighth repetition ($p < 0.05$).

Conclusion: The study was able to establish construct validity for the training program with the simulator under

study. The learning curve showed that training with this simulator is useful for subjects with or without limited endoscopic experience. Furthermore, previous endoscopic camera navigation already improves motor skills to more than the basic level.

Key words: Construct validity — Laparoscopy — Motor skills — Simulator — Training — Virtual reality

Endoscopic virtual reality (VR) trainers have become an attractive and valuable tool for training surgeons in a nonpatient environment. The aim of a simulator is to shorten the learning curve of the trainee for the real job.

The SIMENDO (DelltaTech, Delft, The Netherlands) is a recently developed affordable simulator (5–8 thousand euro for software and hardware) aimed at training hand–eye coordination motor skills needed to perform endoscopic surgery. This simulator is designed to provide an easy-to-use plug-and-play system for surgical trainees. It features abstract tasks and simultaneously measures the performance of subjects using various parameters for objective assessment. However, for practical and effective use in the surgical curriculum, tasks and metrics incorporated in the simulator need to be tested for objectivity and reliability. Consequently, each new training device must be assessed and validated.

In the literature, the validation of training tools and their effectiveness are described using different theoretical models, but exact definitions vary among different authors [2, 5, 11]. One important step in evaluating new training tools is to assess construct validity. Construct validity refers to the concept that the studied novelty (e.g., the SIMENDO VR simulator) measures the quality, ability, or trait it was designed to measure. Therefore, the metrics or parameters assessed must be related to the level of the performer's experience. This usually is accomplished by measuring the performance

of two or more groups that differ in the precise skills being measured by the instrument. For example, practicing endoscopic surgeons should outperform inexperienced trainees. Construct validity has been established for several other VR simulators [3].

In addition to the construct validity study, the learning curves of groups with different experience levels were assessed. This is essential because learning curves determine the training capabilities of the simulator. We propose a theoretical model regarding the requirements of the measured learning curve for the different parameters. A learning curve can be defined as the relationship between the parameters measured through training repetitions.

The model we propose is graphically displayed in Fig. 1. In the model, it is assumed that the parameters such as task time, errors, and path length will decrease as experience is gained. The curve indicated by line 1 characterizes the ideal learning curve for experts using the simulator. Theoretically, there should be no learning effect for experts, because they already poses the ability measured (displayed here as a horizontal line). However, simulation is, per definition, a deduction of a realistic situation. Therefore, in practice, curve 2 represents the expert learning curve more accurately. Curve 3 characterizes the learning curve for novices. Due to novices' lack of experience with the tasks simulated, their curve height is indicative of the performance difference between novices and experts. Obviously, the novices' curve (curve 3) should approach the expert curve in due time. Furthermore, the novices' curve should approach gradually and not too quickly. Tasks mastered within a few repetitions have limited training usefulness. The area between curves 2 and 3 represents the task complexity and the effectiveness of the simulator. Hence, the larger the area between curves 2 and 3 the more difficult is the task or tasks trained in the simulator.

The current study aimed first to establish construct validity by determining whether the parameters measured for the SIMENDO simulator can discriminate groups with various experience levels, and second to establish to what extent training is useful by determining when inexperienced groups reach expert level.

Methods

Participants

The participants were divided into four groups according to their experience with endoscopic surgery:

- *Group A—experienced:* more than 50 endoscopic procedures performed
- *Group B—intermediate:* 1 to 50 endoscopic procedures performed
- *Group C—endoscope navigation:* experienced only in endoscopic camera assistance during endoscopic surgery
- *Group D—novice:* no experience in endoscopic surgery whatsoever.

Materials and techniques

For this study, the SIMENDO virtual reality simulator for endoscopic skills was used. This simulator consists of a software interface with

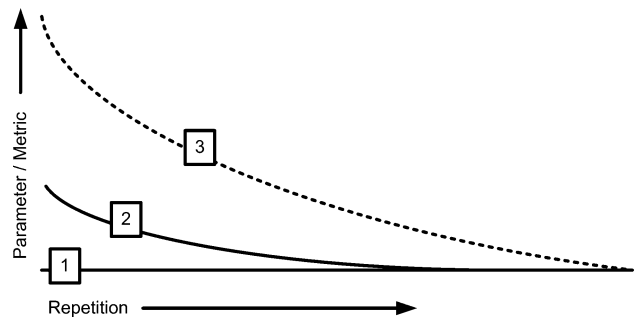


Fig. 1. Theoretical concept for the learning curve of simulator parameter assessment: (1) the theoretical learning curve for experts in an ideal simulation model, (2) the actual measured learning curve for the experts, and (3) the learning curve for novices or nonexperienced subjects.

several training exercises and two hardware instruments. The two instruments were connected with a Universal Serial Bus plug to a standard PC or laptop (Fig. 2). Previously, the usefulness and face validity of this simulator with a single instrument was studied [16].

In the current study, six training exercises from SimSoft (Delltech, Delft, The Netherlands) 1.0 software were included. Table 1 describes the goals and the content of exercises used. In all the exercises, two instruments were used except for the first exercise (drop the balls). Two exercises (drop the balls and stretch) were executed in the same virtual environment with alternating endoscopic camera positions. Impressions of the exercises used in this study corresponding to the descriptions in Table 1 are given in Fig. 2.

All the participants performed three repetitions of the six consecutive exercises described in Table 1. One repetition involved the performance of exercises 1 to 6 without scheduled breaks, except to switch from one exercise to the other. In the second repetition, the participants repeated the same set of exercises from 1 to 6 and so on.

Most of the participants (see detailed description in the Results section) continued training for learning curve assessment and performed a total of 10 repetitions. A break of 5 min was scheduled after the fifth repetition. The training took place in a quiet room in the presence of an observer. The participants received a written instruction that they were not to speak during task execution.

Parameter assessment

The following outcome parameters were automatically generated during performance of the tasks: task time, collisions of instruments with nontarget objects, and total path length for the right and left instruments. Furthermore, the percentage of time the instrument tip was centered in the endoscope and the endoscope path length were measured in two exercises: drop the balls with endoscope navigation and the 30° endoscope. Task time was measured in seconds, collisions in number, and path length in arbitrary units.

The parameters of the six different exercises were summed for each repetition. Then parameter totals per repetition (e.g., total task time, total collisions) were created. After that, the exercises were analyzed individually, in which case, parameters were not summated.

Data were analyzed using the Statistical Package of Social Software (SPSS) version 12.0. The Mann-Whitney *U* test for parametric data was used to analyze statistical differences between the scores of the groups with different levels of experience. Statistical significance was considered when *p* was less than 0.05. Values are presented as median (range) unless stated otherwise.

Results

Construct validation

The 61 participants were divided into four groups by levels of clinical endoscopic experience, as previously defined. This resulted in the following groups:

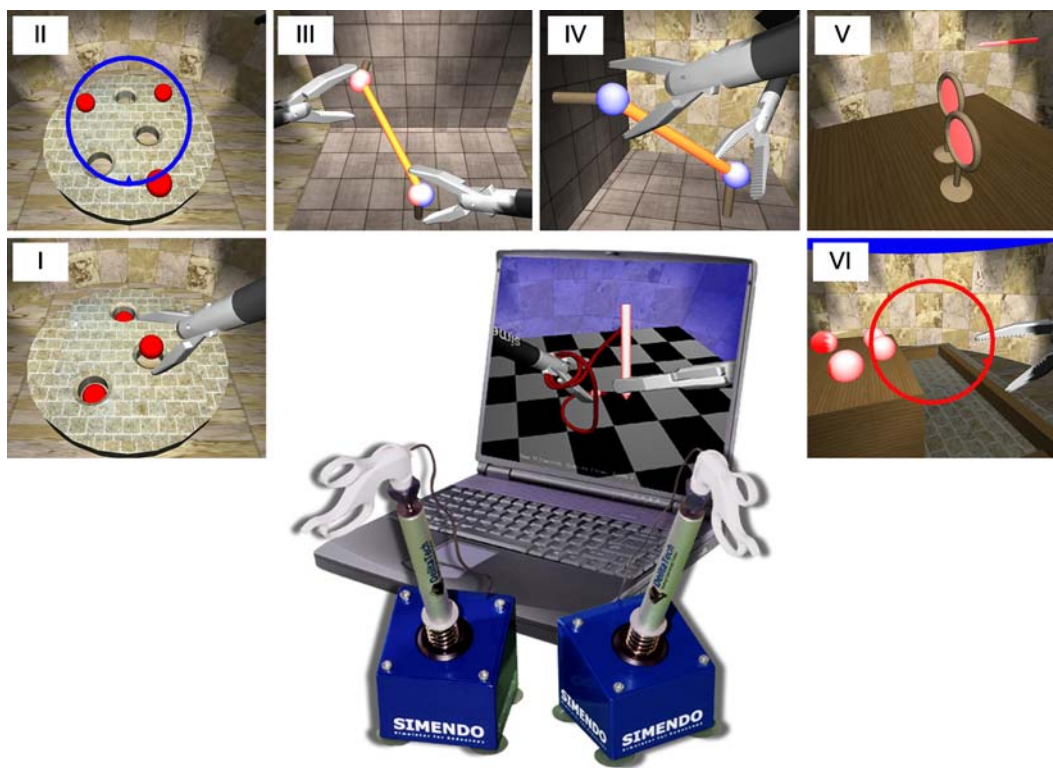


Fig. 2. The SIMENDO simulator for endoscopic surgery. The two instruments are connected to a laptop via USB. The numbers of the six exercises correspond with the description in Table 1.

Table 1. Description of exercises used

Exercise name	Exercise description	Exercise goal
1. Drop the balls, one instrument	Dropping of three balls into holes with right hand only	Basic coordination
2. Drop the balls, two instruments	Dropping of three balls into holes with camera in the left hand and instrument in the right hand	Basic coordination with use of camera
3. Ring and needle	Putting a needle trough 2 rings (both hands)	Fine coordination and positioning
4. Stretch, easy (endoscope 0°)	Stretching a tube in correct direction and length with camera between instruments (both hands)	Easy stretching and coordination
5. Stretch, difficult (endoscope 90°)	Stretching a tube in correct direction and length with camera from the left (90°) (both hands)	Difficult stretching and coordination
6. 30° endoscope handling	Putting 4 balls on a box with a 30° endoscope in the left hand and a grasping instrument in the right hand	Coordination of instrument with 30° camera

- *Group A—experienced* ($n = 15$): median of more than 100 endoscopic procedures performed
- *Group B—intermediates* ($n = 18$): median of 10 (range 1–30) endoscopic procedures performed
- *Group C—endoscope navigation* ($n = 14$): median of 30 (range 1–40) endoscope navigation procedures performed
- *Group D—novices* ($n = 14$): no endoscopic experience.

The results for total task time, number of collisions, endoscope path length and right instrument path length in three consecutive repetitions are given in Fig. 3.

Compared with the expert group, the total task time for the less experienced groups (groups B, C, and D) was significantly longer. The right instrument and endoscope path lengths of the novice and endoscope navigation groups (groups C and D) also were longer. More collisions and a longer left instrument path length in the first two repetitions were found in group D. As compared with the intermediate group (group B), the endoscope and right instrument path lengths were longer in group D. As compared with the endoscope navigation group (group C), longer task times and longer path lengths were found in group D, but this was significant only in the first repetition.

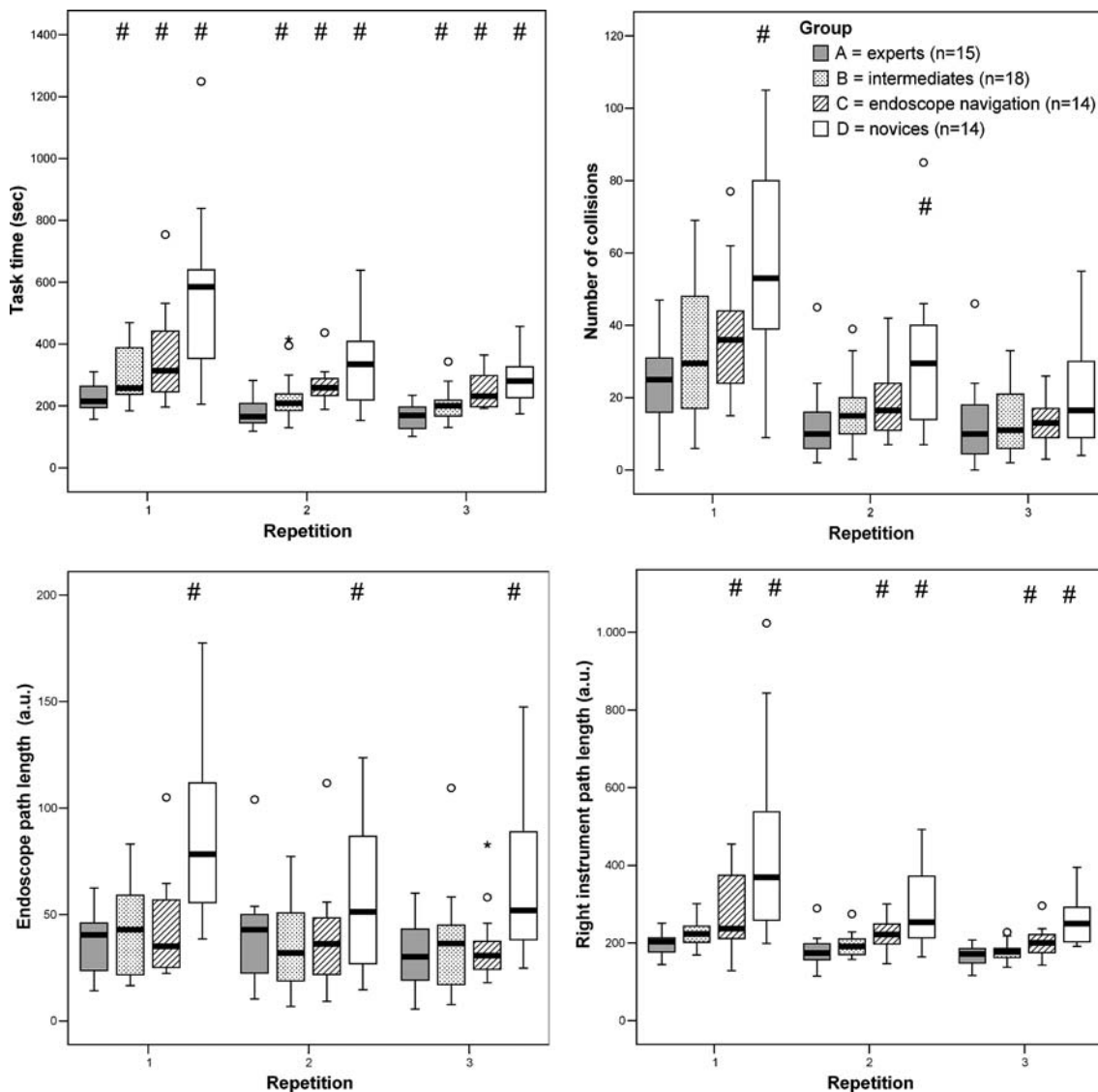


Fig. 3. Box plot diagrams for total task time, collisions, endoscope path length, and right instrument path length for three consecutive training repetitions. The borders of the boxes represent the 25th (lower border) and 75th (upper border) percentiles, and the horizontal line represents the median. The rounds (o) are outliers, defined as 1.5 times

dispersed from the box borders. The stars (*) represent the extreme outliers, defined as 3 times dispersed from the box borders. The symbol (#) above the boxes represents a significant difference ($p < 0.05$) between group B, C, or D and group A (the expert group).

No statistically significant differences were found for the median percentage of time the instrument tip was centered in the endoscope image in the first repetition: expert group (71%) vs intermediate group (70%) vs endoscope navigation group (65%) vs novice group (65%).

Individual tasks

Subsequently, the scores were analyzed for each exercise separately. In this case, the parameters of the different exercises were not summed for each repetition. The task time was significantly longer for the novices (group D) than for the experts (group A) in all six exercises, and it also was longer in three of the six exercises (drop the balls with camera, stretch difficult, and ring and needle)

for the endoscope navigation group (group C). Right instrument path length was longer for the novices (group D) in all tasks except the two stretch exercises, and for the endoscopic navigation group (group C) only in the ring and needle exercise. In addition, the novices had significantly more collisions than the experts (group A) in drop the balls with camera and ring and needle, but not for all repetitions. Finally, there were no significant parameter differences between the experts (group A) and the intermediates (group B) in the individual exercises.

Learning curves

For 49 participants, the learning curve for each parameter was assessed. All the experienced participants (group A,

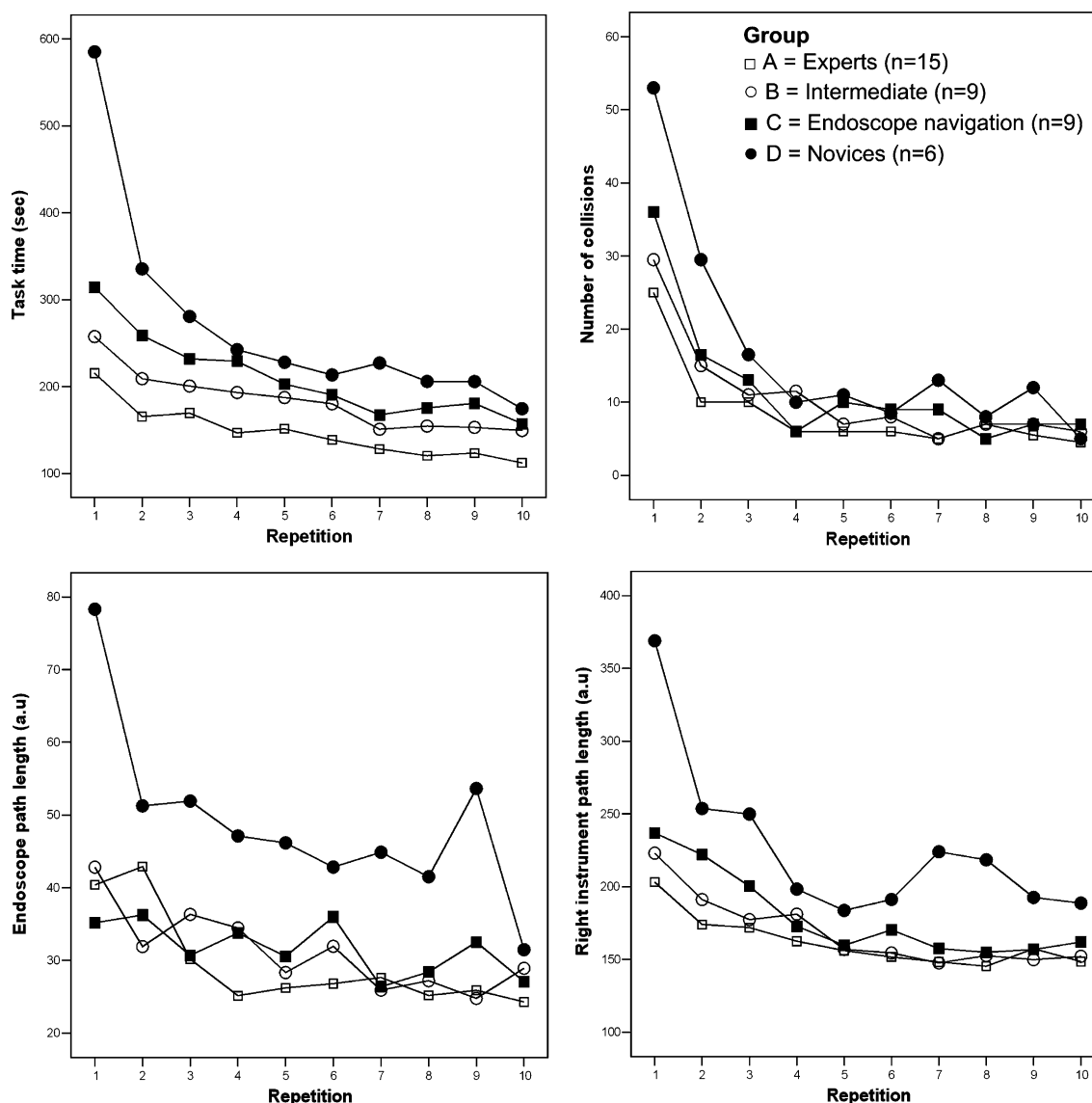


Fig. 4. Learning curves of total median task time, number of collisions, endoscope path length, and right instrument path length for the four groups performing 10 repetitions. The details are given in the text.

$n = 15$) and some of the other groups (intermediate [group B, $n = 9$] endoscope navigation [group C, $n = 9$], and novice [group D, $n = 6$]) continued the training up to 10 repetitions. The results of the learning curve for total task time, right instrument path length, and number of collisions are shown in Fig. 4.

Compared with the expert group, the novice and camera navigation groups had a significantly longer total task time up to the 10th repetition, and the intermediate group up to the 8th repetition. In the novice group, the endoscope and right instrument path lengths also were longer than in the expert group up to the 10th repetition ($p < 0.05$).

Table 2 shows the mean of the parameters for the 1st and 10th repetitions for the experts and novices. Note that the difference is diminished between the expert and novice groups in the 10th repetition. Furthermore, the mean and standard deviation decrease in the novice group from the 1st to the 10th repetition.

Discussion

The first goal of this study was to establish construct validity for the SIMENDO simulator. The results show that the parameters combined with the exercises in the simulator can be related to various levels of laparoscopic experience. Three studied parameters (task time, endoscope path length, and right instrument path length) were able to measure relevant differences over consecutive repetitions for the whole training program and also for some individual tasks. Instrument collisions with the virtual environment as a parameter for error and accuracy can discriminate between novice and experienced subjects only in the beginning of the training (first two repetitions). The learning curve assessment showed that inexperienced groups approached the expert group level by repetitive training. Required practice to achieve expert level was related to previous experience.

Table 2. Expert and novice parameter scores for the 1st and 10th repetitions

Parameter	Repetition	Expert group		Novice group		<i>p</i> Value ^a
		Mean	SD	Mean	SD	
Task time (seconds)	1	230.9	47.3	635.0	364.1	<0.001
	10	118.3	22.9	182.5	61.0	0.032
Collisions (number)	1	25.5	13.3	51.5	13.0	0.010
	10	7.1	6.0	7.7	5.3	NS
Endoscope path length (a.u.)	1	36.9	14.8	98.4	53.8	<0.001
	10	23.3	8.2	42.4	21.9	0.032
Right instrument path length (a.u.)	1	197.1	32.3	468.0	289.5	0.002
	10	144.0	20.2	203.6	52.7	0.005
Left instrument path length (a.u.)	1	65.5	18.3	148.5	82.0	0.003
	10	42.4	7.4	56.7	14.6	NS

SD, standard deviation; NS, not significant; a.u., arbitrary units

^a Two-tailed Mann–Whitney *U* test for expert group vs novice group

The time and path length parameters of the learning curve fit the theoretical model proposed in the introduction (Fig. 1). The learning curve in the expert group for these parameters is lower and more flat, and the difference in relation to the inexperienced groups is great. Analysis of task time along the learning curve shows a significant difference between the groups, even beyond the 10th repetition. The curve shapes indicate that experts adapt rapidly to the tasks, and that inexperienced participants encounter a substantial learning effect.

Strong discriminative ability between levels of experience for time and path length parameters also is observed with other simulators [4, 6, 11, 13–15, 17]. The results in this study suggest that the tasks mimic psychomotor skills needed for endoscopic surgery. In this experiment, none of the inexperienced groups (novice and endoscope navigation groups) fully approach expert level for all parameters. The profound initial difference between the groups and the gradual converging of their learning curves support a substantial training capacity of the device. In contrast to task time and path length, the learning curve for collisions also is short for the inexperienced groups, and therefore does not fit the theoretical model.

Including error assessment in the performance outcome for training devices is imperative because programs that fail to consider objective assessments of accuracy may overestimate endoscopic proficiency [12]. Rapid “flooring” of the error score is also seen in other studies [6]. Apparently, indeed controlled instrument movements avoiding collisions are aptitudes mastered rapidly.

An interesting question is whether the collision parameter in the studied simulator is acceptable as an outcome parameter for accuracy. As a discriminative parameter for experience level, it is not very powerful. However, the SIMENDO aims to train basic dexterity in endoscopic instrument handling (eye–hand coordination) and not in procedural or anatomic rules, instrument–tissue interaction, or knowledge important for tissue handling. Tissue handling requires a high level of accuracy, but for general psychomotor skills training in a basic simulator, a low discriminative power is adequate.

Additionally, inexperienced subjects still needed more time to accomplish the tasks with the same collision number as the experts did. Theoretically, a skilled person is recognized by his or her ability to perform accurately, effectively, and efficiently. In fact, such an observation is clear to anyone who reflects on the difference between their beginner and practiced performance in situations such as learning how to play a musical instrument or learning how to drive a car. Experienced endoscopic surgeons have the ability to combine accuracy (low errors or collisions), effectiveness (short path length), and efficiency (time) well, whereas novices do not yet have that ability. Therefore, the combination of several parameters should be used as criterion-based training goals for inexperienced trainees and not time or path length alone.

Naturally, performance varies among individuals. This variance is represented by the standard deviation in the groups (Table 2). The variance among individuals was greater in the inexperienced novice group than in the experienced group and decreased throughout the training. This observation also has been reported by others [8]. Initially, some inexperienced participants seem to have a more natural dexterity than others in their group. However, in the experienced group, all the participants possess a certain level of dexterity (i.e., a psychomotor skills) developed by their work as endoscopic surgeons.

Despite the large initial differences, attenuation of interindividual performance outcomes occurred (i.e., inexperienced individuals tended to converge to a similar level). According to Ackerman [1], attenuation of variance during task learning indicates a shift from a cognitive conscious process to a more automatic unconscious cognitive process. This is a typical characteristic of motor tasks such as psychomotor skills required in endoscopic surgery. If the trained tasks require predominantly conscious cognitive processes (applying rules, interpreting new situations), only small improvement is seen in the beginning of the curve, but the difference on the interindividual level (e.g., the standard deviation) does not change much throughout training. As expected, the standard deviation did change, which indicates primarily motor skill training in the simulator.

The endoscope navigation group outperformed the novice group by a shorter task time and endoscope path length. Endoscope navigation experience during real laparoscopic procedures improved objective measurable psychomotor skills in the simulator. Despite the substantial number of procedures assisted in this group (median, 30 procedures), quite a bit of additional training was needed to approach expert level in the simulator. It could be assumed that endoscope navigation improves only very specific psychomotor skills, not general dexterity for endoscopic surgery.

Although endoscope navigation may seem to be an easy task, it is, however, of utmost importance because it directly influences the performance of the operating surgeon. Consequently, the simulator also should be used to train novices in specific endoscope navigational skills.

The SIMENDO simulator is designed to train the basic eye–hand coordination necessary for performing endoscopic surgery. The results of this study suggest that the abstract tasks featured in the simulation indeed train the basic motor skills needed to perform endoscopic surgery. These outcomes cannot be related directly to enhanced operative performance, but most likely the SIMENDO simulator will shorten the learning curve of basic psychomotor skills required in the operating room. Training motor skills outside the operating room will facilitate incorporation of these skills on a more automatic level and consequently reduce the mental load during the actual job [7]. The trainee then can concentrate on other aspects of the procedure such as anatomy and procedural steps. As a result, this will increase safety and save expensive operating room training time.

The implication of the learning curve data is that the SIMENDO simulator can be highly effective for novices and surgical trainees with limited endoscopic experience. Its implementation at the beginning of a surgical curriculum for criterion-based licensing of basic psychomotor skills, just before animal training or the first 10 endoscopic procedures, would seem to be most beneficial. Probably, it also would enhance skills maintenance. There is, however, no evidence yet supporting a role for high-stake examination of surgical trainees.

Future research should aim to determine the usefulness of the SIMENDO simulator for recruitment and selection and its position among other validated simulators. Randomized controlled trials with VR simulators such as MIST-VR [9] and LapSim [10] have shown a transfer of skills to the operating room. To determine the position of SIMENDO, the next step should be a comparative study (concurrent validity) with one of these simulators.

Conclusion

This study established construct validity for the SIMENDO. The learning curve showed that the SIMENDO training is useful for subjects with no or

limited endoscopic experience. Furthermore, endoscopic camera navigation during real procedures already improves laparoscopic psychomotor skills to more than the basic level.

Acknowledgment. This research was financially supported by a grant provided by the health insurance company DSW (Schiedam, The Netherlands).

References

1. Ackerman PL (1987) Individual differences in skill learning: an integration of psychometric and information processing perspectives. *Psychol Bull* 102: 3–27
2. Aggarwal R, Moorthy K, Darzi A (2004) Laparoscopic skills training and assessment. *Br J Surg* 91: 1549–1558
3. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Jakimowicz JJ (2005) Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc* 19: 1523–1532
4. Duffy AJ, Hogle NJ, McCarthy H, Lew JI, Egan A, Christos P, Fowler DL (2004) Construct validity for the LAPSIM laparoscopic surgical simulator. *Surg Endosc* 19(3): 401–405
5. Feldman LS, Hagarty SE, Ghitulescu G, Stanbridge D, Fried GM (2004) Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. *J Am Coll Surg* 198: 105–110
6. Gallagher AG, Lederman AB, McGlade K, Satava RM, Smith CD (2004) Discriminative validity of the Minimally Invasive Surgical Trainer in Virtual Reality (MIST-VR) using criteria levels based on expert performance. *Surg Endosc* 18: 660–665
7. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, Smith CD, Satava RM (2005) Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 241: 364–372
8. Gallagher AG, Satava RM (2002) Virtual reality as a metric for the assessment of laparoscopic psychomotor skills: learning curves and reliability measures. *Surg Endosc* 16: 1746–1752
9. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P (2004) Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg* 91: 146–150
10. Hyltander A, Liljegren E, Rhodin PH, Lonroth H (2002) The transfer of basic skills learned in a laparoscopic simulator to the operating room. *Surg Endosc* 16: 1324–1328
11. Schijven M, Jakimowicz J (2003) Construct validity: experts and novices performing on the Xitact LS500 laparoscopy simulator. *Surg Endosc* 17: 803–810
12. Smith CD, Farrell TM, McNatt SS, Metreveli RE (2001) Assessing laparoscopic manipulative skills. *Am J Surg* 181: 547–550
13. Smith SG, Torkington J, Brown TJ, Taffinder NJ, Darzi A (2002) Motion analysis. *Surg Endosc* 16: 640–645
14. Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A (1998) Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST-VR laparoscopic simulator. *Stud Health Technol Inform* 50: 124–130
15. Van Sickle KR, McClusky DA III, Gallagher AG, Smith CD (2005) Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. *Surg Endosc* 19: 1227–1231
16. Verdaasdonk EG, Stassen LP, Monteny LJ, Dankelman J (2006) Validation of a new basic virtual reality simulator for training of basic endoscopic skills: the SIMENDO. *Surg Endosc* 20: 511–518
17. Woodrum DT, Andreatta PB, Yellamanchilli RK, Feryus L, Gauger PG, Minter RM (2006) Construct validity of the LapSim laparoscopic surgical simulator. *Am J Surg* 191: 28–32